On the derivation of the Dirac equation

R.Sartor*

University of Liege, Sart-Tilman, Institute of Physics B5, B-4000 Liege 1,Belgium

(Dated: March 23, 2009)

Abstract

We point out that the anticommutation properties of the Dirac matrices can be derived without squaring the Dirac hamiltonian, that is, without any explicit reference to the Klein-Gordon equation. We only require the Dirac equation to admit two linearly independent plane wave solutions with positive energy for all momenta. The necessity of negative energies as well as the trace and determinant properties of the Dirac matrices are also a direct consequence of this simple and minimal requirement.

I. INTRODUCTION

Many textbooks[1,2,3,4,5,6] derive the Dirac equation for a free particle of mass m following the method used by Dirac himself in his 1928 paper[7]. This method involves two steps. First, one admits that the wave function should be a multi-component object, as in the non-relativistic theory of spin, and that its time evolution is ruled by a partial differential equation of first order in both the time and space derivatives. Working in a system of units where $\hbar = c = 1$, we have

$$i\frac{\partial \Psi}{\partial t} = H_D \Psi, \tag{1}$$

where the Dirac hamiltonian H_D is defined by

$$\begin{aligned} H_D &= \boldsymbol{\alpha} \cdot (-i\nabla) + \beta m \\ &= \sum_{k=1}^{3} \alpha_k (-i\frac{\partial}{\partial x^k}) + \beta m. \end{aligned} \tag{2}$$

In this equation, the α_k's and β are constant hermitian matrices. In the second step, one 'squares' Eq. (1) by acting on both sides of it with the operator $i\frac{\partial}{\partial t}$. This yields

$$\begin{aligned} -\frac{\partial^2 \Psi}{\partial t^2} &= H_D(i\frac{\partial}{\partial t}\Psi) \\ &= H_D^2 \Psi. \end{aligned} \tag{3}$$

Then one requires H_D^2 to be identical to the operator $-\Delta + m^2$, thereby ensuring that each of the components of Ψ satisfies the Klein-Gordon equation.This implies the anticommutation relations

$$\begin{aligned} \alpha_i \alpha_j + \alpha_j \alpha_i &= 2\delta_{ij}, \\ \alpha_i \beta + \beta \alpha_i &= 0, \\ \beta^2 = 1. \end{aligned} \tag{4}$$

Starting from these results, one usually proceeds by showing that such matrices indeed exist when Ψ is a four-component object and then one 'finds' that Eq. (1) admits both positive and negative energy plane wave solutions. So the Dirac equation does not solve the 'problem of negative energies' which appears when studying the Klein-Gordon equation.

However, it is difficult to be immediately convinced that this fact is not a mere consequence of the requirement appearing in the second step of the above derivation. Thus, to

discard any doubts about the necessity of negative energies, it would be more satisfactory to avoid the squaring of the hamiltonian H_D. In this paper, we show that this is indeed feasible.

II. NECESSITY OF NEGATIVE ENERGIES

In order to implement the program outlined at the end of the previous section, we require Eq. (1) to admit two linearly independent plane wave solutions of the form

$$\Psi(\mathbf{x}, t) = u(\mathbf{p}) e^{i(\mathbf{p}\cdot\mathbf{x} - E_{\mathbf{p}}\, t)}, \tag{5}$$

where $E_{\mathbf{p}}$ is the positive energy associated with a free particle of momentum $\mathbf{p}$, that is,

$$E_{\mathbf{p}} = +(\mathbf{p}^2 + m^2)^{1/2}. \tag{6}$$

Since our aim is to describe spin 1/2 particles such as electrons, this requirement is both natural and minimal. By inserting Eq. (5) into Eq. (1), we obtain

$$E_{\mathbf{p}}\, u(\mathbf{p}) = h_D(\mathbf{p})\, u(\mathbf{p}), \tag{7}$$

with

$$h_D(\mathbf{p}) = \boldsymbol{\alpha}\cdot\mathbf{p} + \beta m. \tag{8}$$

Note that $h_D(\mathbf{p})$ is a matrix of numbers whereas H_D is a matrix of differential operators. Since we obviously discard the solution $u(\mathbf{p}) = 0$, we see, from the above requirement, that $E_{\mathbf{p}}$ should be a double root of the eigenvalue equation pertaining to the matrix $h_D(\mathbf{p})$. Thus, if we introduce the characteristic polynomial of $h_D(\mathbf{p})$

$$P_n(E) = dtm[E - h_D(\mathbf{p})], \tag{9}$$

we should have

$$P_n(E_{\mathbf{p}}) = 0, \tag{10}$$

and

$$P_n'(E_{\mathbf{p}}) = 0, \tag{11}$$

where P_n' is the derivative of P_n with respect to E. The index n in these equations stands for the degree of $P_n(E)$ or, equivalently, for the number of components of the wave function Ψ.

Let us now try to satisfy Eqs. (10) and (11) within a two-component theory ($n = 2$). We have

$$P_n(E) \equiv P_2(E) = E^2 + c_1(\mathbf{p})E + c_0(\mathbf{p}), \tag{12}$$

where the coefficients c_1 and c_0 are polynomials homogeneous in m and the components p_1, p_2, p_3 of the momentum $\mathbf{p}$. Eq. (11) yields

$$2E_{\mathbf{p}} + c_1(\mathbf{p}) = 0. \tag{13}$$

It is not possible to satisfy this equation for all momenta since the square root $E_{\mathbf{p}}$ cannot be expressed as a polynomial. Thus, a two-component theory is immediately ruled out. So, let us try a Dirac equation with three components. Now, we have

$$P_n(E) \equiv P_3(E) = E^3 + c_2(\mathbf{p})E^2 + c_1(\mathbf{p})E + c_0(\mathbf{p}), \tag{14}$$

where the coefficients c_2, c_1 and c_0 are again polynomials homogeneous in m and the components of the momentum $\mathbf{p}$. Eqs. (10) and (11) yield

$$E_{\mathbf{p}}^3 + c_2(\mathbf{p})E_{\mathbf{p}}^2 + c_1(\mathbf{p})E_{\mathbf{p}} + c_0(\mathbf{p}) = 0, \tag{15}$$

$$3E_{\mathbf{p}}^2 + 2c_2(\mathbf{p})E_{\mathbf{p}} + c_1(\mathbf{p}) = 0. \tag{16}$$

Again using the fact that $E_{\mathbf{p}}$ cannot be expressed as a polynomial, we see that these equations imply

$$E_{\mathbf{p}}^2 + c_1(\mathbf{p}) = 0, \tag{17}$$

$$c_2(\mathbf{p})E_{\mathbf{p}}^2 + c_0(\mathbf{p}) = 0, \tag{18}$$

$$3E_{\mathbf{p}}^2 + c_1(\mathbf{p}) = 0, \tag{19}$$

$$c_2(\mathbf{p}) = 0. \tag{20}$$

Eqs. (17) and (19) lead to $E_{\mathbf{p}} = 0$ for all momenta. This is not possible and, as a consequence, a three-component Dirac theory is also ruled out. Finally, let us turn to a four-component theory. Now,

$$P_n(E) \equiv P_4(E) = E^4 + c_3(\mathbf{p})E^3 + c_2(\mathbf{p})E^2 + c_1(\mathbf{p})E + c_0(\mathbf{p}), \tag{21}$$

where our notations are similar to those used above in the two- and three-component cases. Eqs. (10) and (11) yield

$$E_{\mathbf{p}}^4 + c_3(\mathbf{p})E_{\mathbf{p}}^3 + c_2(\mathbf{p})E_{\mathbf{p}}^2 + c_1(\mathbf{p})E_{\mathbf{p}} + c_0(\mathbf{p}) = 0, \tag{22}$$

$$4E_{\mathbf{p}}^3 + 3c_3(\mathbf{p})E_{\mathbf{p}}^2 + 2c_2(\mathbf{p})E_{\mathbf{p}} + c_1(\mathbf{p}) = 0. \tag{23}$$

These equations imply

$$E_{\mathbf{p}}^4 + c_2(\mathbf{p})E_{\mathbf{p}}^2 + c_0(\mathbf{p}) = 0, \tag{24}$$

$$c_3(\mathbf{p})E_{\mathbf{p}}^2 + c_1(\mathbf{p}) = 0, \tag{25}$$

$$2E_{\mathbf{p}}^2 + c_2(\mathbf{p}) = 0, \tag{26}$$

$$3c_3(\mathbf{p})E_{\mathbf{p}}^2 + c_1(\mathbf{p}) = 0. \tag{27}$$

From Eq. (26), we obtain

$$c_2(\mathbf{p}) = -2E_{\mathbf{p}}^2. \tag{28}$$

Inserting this expression into Eq. (24) yields

$$c_0(\mathbf{p}) = E_{\mathbf{p}}^4. \tag{29}$$

Finally, comparing Eqs. (25) and (27) leads to

$$c_1(\mathbf{p}) = 0 \tag{30}$$

and

$$c_3(\mathbf{p}) = 0. \tag{31}$$

If we insert these results back into Eq. (21), we see that the eigenvalue equation for $h_D(\mathbf{p})$ reads

$$(E - E_{\mathbf{p}})^2(E + E_{\mathbf{p}})^2 = 0. \tag{32}$$

This shows that the positive energy solutions to the Dirac equation will always be accompanied by solutions with negative energy. To prove that the approach adopted in this paper is self-contained, we still have to derive the anticommutation relations (4). This is performed in the next section.

III. DERIVATION OF THE ANTICOMMUTATION RELATIONS

We now show that Eqs. (28), (29), (30) and (31) do indeed imply Eqs. (4). We remark that once we have replaced $E_{\mathbf{p}}$ by its expression (6), all of these equations require some polynomial homogeneous in m and the components of $\mathbf{p}$ to vanish identically, that is for

all momenta. This is possible only if all the polynomial coefficients are zero. We shall rely repeatedly on this remark in what follows.

From Eqs. (9) and (21), we obtain

$$\begin{aligned} c_3(\mathbf{p}) &= -Tr(h_D(\mathbf{p})) \\ &= \sum_{k=1}^{3} p_k Tr(\alpha_k) + mTr(\beta), \end{aligned} \tag{33}$$

where the symbol Tr denotes the trace. Thus, Eq.(31) implies

$$Tr(\alpha_1) = Tr(\alpha_2) = Tr(\alpha_3) = Tr(\beta) = 0. \tag{34}$$

Eqs.(9) and (21) also yield

$$c_0(\mathbf{p}) = dtm(h_D(\mathbf{p})). \tag{35}$$

Inserting this expression into Eq.(29) and considering the terms in p_1^4, p_2^4, p_3^4 and m^4 leads to

$$dtm(\alpha_1) = dtm(\alpha_2) = dtm(\alpha_3) = dtm(\beta) = 1. \tag{36}$$

To make things simpler, it is convenient to work in a representation where the matrix β is diagonal. Note that Eqs.(34) and (36) are representation independent. Consider the terms in m^3 in Eq.(30) and in m^2 in Eq.(28). They yield

$$\beta_{11}\beta_{22}\beta_{33} + \beta_{11}\beta_{22}\beta_{44} + \beta_{11}\beta_{33}\beta_{44} + \beta_{22}\beta_{33}\beta_{44} = 0 \tag{37}$$

and

$$\beta_{11}\beta_{22} + \beta_{11}\beta_{33} + \beta_{11}\beta_{44} + \beta_{22}\beta_{33} + \beta_{22}\beta_{44} + \beta_{33}\beta_{44} = -2, \tag{38}$$

respectively. Combining these equations with

$$dtm(\beta) = \beta_{11}\beta_{22}\beta_{33}\beta_{44} = 1, \tag{39}$$

(see Eq.(36)), we obtain

$$(1 + \beta_{11})(1 + \beta_{22})(1 + \beta_{33})(1 + \beta_{44}) = 0 \tag{40}$$

and

$$(1 - \beta_{11})(1 - \beta_{22})(1 - \beta_{33})(1 - \beta_{44}) = 0. \tag{41}$$

These equations show that one of the eigenvalues of β is equal to $+1$ and another to -1. Let us assume that $\beta_{11} = +1$ and $\beta_{33} = -1$. Taking Eqs.(34) and (36) into account, this

implies $\beta_{44} = -\beta_{22}$ and $\beta_{22} = \pm 1$. We shall assume that $\beta_{22} = +1$ and thus $\beta_{44} = -1$. We do not have to consider other choices for the diagonal elements to be put equal to $+1$ or -1 since this would correspond to a mere rearrangement of the lines and columns of β. Thus, we have

$$\beta = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}. \tag{42}$$

Obviously,

$$\beta^2 = 1, \tag{43}$$

and it is easy to show that this equation implies

$$\alpha_i^2 = 1 \quad (i = 1, 2, 3). \tag{44}$$

Indeed, let us just imagine that we perform, on all the Dirac matrices, a unitary transformation which brings α_1, say, into diagonal form. We expect that the matrix β will no longer be diagonal but Eq.(43) will remain true because it is representation independent. We now proceed for α_1 as we did above for β, that is, we concentate on the terms in p_1^3 in Eq.(30) and in p_1^2 in Eq.(28). This will lead us to $\alpha_i^2 = 1$ which is also representation independent. Proceeding in this way for α_2 and α_3, we prove the other identities in Eq.(44). This trick can be used each time we establish a representation independent identity even if we arrived at that identity within a particular representation. In what follows, we go back to the representation in which β is given by Eq.(42) and we derive the structure of α_1 in that representation. For the moment, we drop the index 1 to simplify our notations. Thus, α stands for α_1. Consider Eq. (28). The terms in mp_1 and in p_1^2 give

$$-\alpha_{11} - \alpha_{22} + \alpha_{33} + \alpha_{44} = 0 \tag{45}$$

and

$$\begin{aligned} &\alpha_{11}\alpha_{22} + \alpha_{11}\alpha_{33} + \alpha_{11}\alpha_{44} + \alpha_{22}\alpha_{33} + \alpha_{22}\alpha_{44} + \alpha_{33}\alpha_{44} \\ &-|\alpha_{12}|^2 - |\alpha_{13}|^2 - |\alpha_{14}|^2 - |\alpha_{23}|^2 - |\alpha_{24}|^2 - |\alpha_{34}|^2 = -2, \end{aligned} \tag{46}$$

respectively. Consider now the terms in $m^2 p_1^2$ in Eq.(29), they give

$$\begin{aligned} &\alpha_{11}\alpha_{22} - \alpha_{11}\alpha_{33} - \alpha_{11}\alpha_{44} - \alpha_{22}\alpha_{33} - \alpha_{22}\alpha_{44} + \alpha_{33}\alpha_{44} \\ &-|\alpha_{12}|^2 + |\alpha_{13}|^2 + |\alpha_{14}|^2 + |\alpha_{23}|^2 + |\alpha_{24}|^2 - |\alpha_{34}|^2 = 2. \end{aligned} \tag{47}$$

Adding Eqs.(46) and (47) yields

$$\alpha_{11}\alpha_{22} + \alpha_{33}\alpha_{44} - |\alpha_{12}|^2 - |\alpha_{34}|^2 = 0. \tag{48}$$

On the other hand, comparing Eq.(45) with Eq.(34) yields

$$\alpha_{22} = -\alpha_{11} \tag{49}$$

and

$$\alpha_{44} = -\alpha_{33}. \tag{50}$$

If we insert these results back into Eq.(48), we obtain

$$\alpha_{11}^2 + \alpha_{33}^2 + |\alpha_{12}|^2 + |\alpha_{34}|^2 = 0. \tag{51}$$

Thus, we have

$$\alpha_{11} = \alpha_{22} = \alpha_{33} = \alpha_{44} = \alpha_{12} = \alpha_{34} = 0, \tag{52}$$

and, restoring the index 1, we see that the matrix α_1 has the following structure:

$$\alpha_1 = \begin{pmatrix} 0 & 0 & (\alpha_1)_{13} & (\alpha_1)_{14} \\ 0 & 0 & (\alpha_1)_{23} & (\alpha_1)_{24} \\ (\alpha_1)_{13}^* & (\alpha_1)_{23}^* & 0 & 0 \\ (\alpha_1)_{14}^* & (\alpha_1)_{24}^* & 0 & 0 \end{pmatrix}, \tag{53}$$

where the non-vanishing elements are restricted by the condition

$$|(\alpha_1)_{13}|^2 + |(\alpha_1)_{14}|^2 + |(\alpha_1)_{23}|^2 + |(\alpha_1)_{24}|^2 = 2. \tag{54}$$

Actually, this equation tells us nothing new since it can be derived from Eq.(44). An analogous proof shows that the matrices α_2 and α_3 have also this structure. We note that the structure of the α_i's and of β (see Eq.(42)) imply

$$\alpha_i\beta + \beta\alpha_i = 0 \quad (i = 1, 2, 3), \tag{55}$$

as can be checked simply by performing matrix multiplications. Since these equations are representation independent, we conclude, using the trick described after Eq.(44), that we should also require

$$\alpha_i\alpha_j + \alpha_j\alpha_i = 0 \quad (i \neq j = 1, 2, 3). \tag{56}$$

Eqs.(43), (44), (55) and (56) are the anticommutation relations we were looking for. It is easy to check that Eqs.(28), (29) and (30) do not give rise to additional restrictions on the Dirac matrices. As an example, consider the terms in p_1p_2 in Eq.(28). They impose

$$\begin{aligned} &(\alpha_1)^*_{13}(\alpha_2)_{13} + (\alpha_1)_{13}(\alpha_2)^*_{13} + (\alpha_1)^*_{14}(\alpha_2)_{14} + (\alpha_1)_{14}(\alpha_2)^*_{14} \\ + \ &(\alpha_1)^*_{23}(\alpha_2)_{23} + (\alpha_1)_{23}(\alpha_2)^*_{23} + (\alpha_1)^*_{24}(\alpha_2)_{24} + (\alpha_1)_{24}(\alpha_2)^*_{24} = 0. \end{aligned} \tag{57}$$

This equation can be written as

$$(\alpha_1\alpha_2 + \alpha_2\alpha_1)_{11} + (\alpha_1\alpha_2 + \alpha_2\alpha_1)_{22} = 0 \tag{58}$$

and, indeed, requires nothing new.

IV. SUMMARY AND COMMENTS

In this paper, we have provided a method to derive the anticommutation properties of the Dirac matrices without relying on the squaring of the Dirac hamiltonian. We have only required Eq.(1) to admit two linearly independent plane wave solutions with positive energy for all momenta. At an early stage in the derivation, we have seen that, despite this conservative requirement, it was not possible to rule out negative energy solutions, thereby establishing that these are not an artefact of the standard derivation. It might also be interesting to note that, within the method described in this paper, the trace and determinant properties of the Dirac matrices appear in the course of the derivation and not as by-products of the anticommutation relations. Finally, a few comments are appropriate concerning our proof of the impossibility of a two-component Dirac equation. As is well known, such an equation appears in a space-time with less than three space dimensions[6] or in the study of massless fermions, where it is known as the Weyl equation[1,4]. This by no means contradicts our assertions. Indeed, no impossibility arises in a two-component theory if one only requires the Dirac equation to admit a single plane wave solution with positive energy for all momenta. In that case, only Eq.(10) with $n = 2$ has to be imposed and this yields

$$E^2_{\mathbf{p}} + c_1(\mathbf{p})E_{\mathbf{p}} + c_0(\mathbf{p}) = 0. \tag{59}$$

This equation implies

$$E^2_{\mathbf{p}} + c_0(\mathbf{p}) = 0, \tag{60}$$

and

$$c_1(\mathbf{p}) = 0. \tag{61}$$

From these equations, we see that the eigenvalue equation for $h_D(\mathbf{p})$ reads

$$(E - E_{\mathbf{p}})(E + E_{\mathbf{p}}) = 0. \tag{62}$$

Thus, we have a plane wave solution with positive energy and another with negative energy. As a consequence, in a two-component theory, the 'twofold degeneracy' only corresponds to the existence of antiparticles. The derivation of the properties of the Dirac matrices (actually, of the Pauli matrices, since we are now in a two-component theory) can be performed as in the previous section and will not be repeated here.

---

[*] R.Sartor@ulg.ac.be

[1] J. D. Bjorken and S. D. Drell, *Relativistic Quantum Mechanics* (Mc Graw-Hill, NY, 1964).

[2] A. Messiah, *Quantum Mechanics* (Dover Publications, Mineola, NY, 2000)

[3] P. Roman, *Advanced Quantum Theory* (Addison Wesley, Reading, MA, 1965)

[4] M. D. Scadron, *Advanced Quantum Theory* (Springer Verlag, NY, 1979)

[5] L. I. Schiff, *Quantum Mechanics* (Mc Graw-Hill Kogakusha, Tokyo, 1968), 3rd. ed.

[6] B. Thaller, *Advanced Visual Quantum Mechanics* (Springer Verlag, NY, 2005)

[7] P. A. M. Dirac, "The Quantum Theory of the Electron," Roy. Soc. Proc. **A117**, 610-624 (1928)